|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Nominal |
| Weight of a person | Ratio |
| Weight of Gold | Ratio |
| Distance between two places | Ratio |
| Length of a leaf | Ratio |
| Dog's weight | Ratio |
| Blue Color | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Ordinal |
| Number of times married | Discrete |
| Gender (Male or Female) | Nominal |

Q1) Identify the Data type for the Following:

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal |
| High School Class Ranking | Ordinal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinl |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Interval |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

Possible events with equal probability :

HHH ,HHT, HTH ,HTT, THH, THT, TTH, TTT

Number with 2 heads: 3

Total number: 8

probability, the number is 3/8 = 0.375 = 37.5%

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2and 3

Ans- a) there is no outcome the sum is equal to 1 . probablity is 0.

b)(1,3)(2,2),(3,1)=3 outcome = 3/36 i.e 1/12

c) 6/36= 1/6

Q5) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

Ans- 10/21

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of count of candies for children(ignoring the nature of the child-Generalized view)

|  |  |  |
| --- | --- | --- |
| CHILD | Candies count | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

Ans=  1 \* 0.015  + 4\*0.20  + 3 \*0.65  + 5\*0.005  + 6 \*0.01  + 2 \* 0.12

= 0.015 + 0.8  + 1.95 + 0.025 + 0.06 + 0.24

=       3.090

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | mean | median | mode | variance | Standard deviation |
| point | 3.59 | 3.695 | 3.92 | 0.28588 | 0.5346 |
| score | 2.21 | 3.325 | 3.44 | 0.9573 | 0.9784 |
| weight | 17.84 | 17.71 | 17.02 | 3.1931 | 1.786 |

Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

Ans= 145.33

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans**

**Skewness for speed = -0.1139548, skewness value is negative so it is left skewed. since magnitude is slightly greater rthan 0 it is slightly left skewed.**

**for disteance = 0.7824835, right skewed(+ve) slight magnitude to right.**

**Q10) Draw inferences about the following boxplot & histogram**



Ans=

the most of the data points are concerted in the range 50-100 with frequency 200

and least range of weight is 400 somwre around 0-10.

so the expected value the above distribution is 75.

skewness- we can notice a long tail towards right so its is heavily right skew

**Q11)**Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

Ans

X=/-(Z1-α.σ/sqrt(n)

Degree of freedom = 2000-1=1999

confidence interval = 94%

(1-σ/2= 1-0.03)= 0.97

for confidence interval for 94% is 1.882

confidence interval for 98% = 2.33

confidence interval for 96% = 2.05

**Q12)**Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean,median,variance,standard deviation.
2. What can we say about the student marks?

Ans=

Mean=41, Median=40, Variance=24.111, standard deviation=4.910

Q13) What is the nature of skewness when mean, median of data are equal?

Ans= Symmetrical

Q14) What is the nature of skewness when mean >median ?

Ans= Right skewed

Q15) What is the nature of skewness when median > mean?

Ans= Left skewed

Q16) What does positive kurtosis value indicates for adata ?

Ans= the data is normally distributed and kurtosis value is 0

Q17) What does negative kurtosis value indicates for a data?

Ans = the distribution of data is lighter tails and a flatter peaks than the normal distribution

Q18) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

What is nature of skewness of the data?

What will be the IQR of the data (approximately)?   
  
Ans=

\*Distribution data=

lets assume above the box plot is about age's of the student in a school.

50% of the people are above 10 years old and remaining are less.

and students who's aage is above 15 are approx 40%

\*Nature of skewness= left skewed median is greater than mean

\*IQR of data is approximately= -8

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

Ans= mean and median are equal hence distribution is symmetrical.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

a) there are 33 observation in MPG are greater than 38

b) there are 61 observation which are lesser than 40

c) P (20<MPG<50)= 69

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans= it follows normality by the w value=0.97797andp value=0.1764which is greater than 0.05

1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans= For WC it follows normality as the mean value (91.902) is not farfrom median (90.8).For AT it is not normally distributed as the p value = 0.000654 which islower than 0.05 and mean(101.894) and median(96.54) values are fardifferent

Q 22) Calculate the Z scoresof 90% confidence interval,94% confidence interval, 60% confidence interval

Ans=

z score of 90% =1.65

Z score of 94% = 1.56

Z score of 60% = 0.85

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

For 95%= 1.96

For 96%= 2.5

For 99% = 2.47

Q 24**)**A Government companyclaims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode🡪pt(tscore,df)

df 🡪 degrees of freedom

Ans=

x = mean of the sample of bulbs =  260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

t = - 0.471

For probability calculations, the number of degrees of freedom is n - 1, so here you need the t-distribution with 17 degrees of freedom.

The probability that t < - 0.471 with 17 degrees of freedom assuming the population mean is true, the t-value is less than the t-value obtained With 17 degrees of freedom and a t score of - 0.471, the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulbs is 300 days.